DOCUMENT RESUME

ED 260 382 CS 008 127

AUTHOR Gutkin, Joan
TITLE The Effect of Diagnostic Inservice Training on Class Reading Achievement and the Number of Lessons Covered.
PUB DATE 85
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985). Parts of document contain small print.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Inservice Teacher Education; Language Arts; Pretests Posttests; *Reading Achievement; *Reading Diagnosis; *Reading Research; Reading Skills; Reading Tests; Socioeconomic Status; Student Evaluation

ABSTRACT

A study was conducted to determine whether a particular diagnostic assessment and monitoring program would have an effect on class achievement, regardless of curriculum used. The program consisted of a series of criterion-referenced tests coordinated with sequenced lessons in reading, oral language, and arithmetic, with the results charted by skill and pupil. These results formed a basis for ongoing assistance to the teacher by an inservice teacher trainer or supervisor. Conducted in an urban school district with poor pupil achievement, heavy bilingualism, and low pupil socioeconomic status, the study used an experimental pretest posttest comparison group design. All teachers were instructed to use the district's specific language program daily. There were three subtests--phonics, reading vocabulary, and reading comprehension--and a total reading score. Although the comparison group began with higher pretest results in every subtest, the experimental group scored higher in every area except reading vocabulary, and made greater gains in every subtest. However, none of the results reached significance. For the experimental group, covering more lessons in the language program, together with the diagnostic inservice training, resulted in a significant positive correlation between lesson day in program and the posttest scores on every subtest. The more the experimental group teachers covered in the language program, the higher the resulting scores in all areas. (Tables of findings are included.) (HTH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED260382

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

BEST COPY AVAILABLE

The Effect of Diagnostic Inservice Training on Class Reading Achievement and the Number of Lessons Covered

Joan Gutkin, Ph. D.
Community School District 1
80 Montgomery Street
New York, New York 10002

New York University

Paper presented at the Annual Meeting
American Educational Research Association
Chicago, Illinois
March 31-April 4, 1985

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Joan Gutkin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS 008 127

Many teachers feel frustrated and helpless in their attempts to improve instruction, particularly with students from lower socio-economic levels. However, inservice staff development generally consists of group sessions, sometimes conducted by outside experts, which by its nature cannot deal with specific problems in specific classrooms. The supervisory observations that occur periodically are more likely to concentrate on the analysis of a particular lesson rather than on an evaluation of the whole instructional sequence (Acheson & Gall, 1980).

Therefore, the teacher in the elementary school receives very little diagnostic feedback on pupil performance during the year. It comes only at the end of the year when it is too late to remediate. This is particularly true in the critical first grade. The main responsibility for ongoing diagnosis, test construction, instruction and remediation lies with the individual classroom teacher, who can expect to receive very little specific guidance or feedback from supervisors in these areas.

The present investigation grew out of the writer's previous experience with the Direct Instruction Follow Through Project at P.S. 137, Brooklyn. That program uses a series of criterion -referenced tests coordinated with its numerically sequenced lessons in reading, oral language, and arithmetic, and the results are charted by skill and pupil. These results form a basis for ongoing assistance to the teacher by a teacher trainer or supervisor, based on the particular lesson skills missed on the tests (Meyer, Gersten, & Gutkin, 1983).

Although that program utilizes diagnostic inservice training, no study has been conducted to compare the effectiveness of its curriculum directly, by measuring it with and without this monitoring of student progress and its concommitant staff development activities to see whether the success could be attributed to the curriculum or to the inservice training. The present research applied the same monitoring and correctional techniques to other curricular materials in other educational settings to see if the assessment and monitoring system has an effect on class achievement, regardless of curriculum used.

The diagnostic inservice training model used in this study integrates concepts from Mastery Learning (Bloom, 1976), Coaching (Joyce & Showers, 1981), process-product teacher research (Mitzel, cited in Rosenshine & Furst, 1971), and performance-based teacher education (PBTE) (Gage & Winne, 1975).

The concept of monitoring student achievement through periodic testing and the use of those test results as a basis for improved instruction is derived from mastery learning. However, mastery learning does not incorporate the periodic test results into its staff development training. There is no provision for skill training. Mastery learning assumes that teachers are capable of the remediation themselves once the pupil test results are known.

The mastery learning model concepts of periodic testing and improved instruction through the use of these test results, which make teachers aware of student achievement on an ongoing basis, became part of the model used in diagnostic inservice training. However, as the feedback from the mastery learning testing is used by the teachers themselves, without mediation by others, another component was needed.

An additional concept, training teachers in skills that are directly tied to specifiable student performance in an evaluation-feedback-training cycle, was derived from PBTE. The PBTE movement defines teacher training as that in which the teacher acquires performance tendencies and capabilities that promote student achievement to a specified level of mastery by the teacher.

Coaching extends the concepts even further by adding hands-on assistance to the teacher, in the classroom, on an ongoing basis, to modify teaching strategies. It is used when the knowledge of the results do not provide sufficient capability for change.

The contribution of using student achievement results (one of the outcome measures) as a criterion of teacher performance is derived from studies of effective teaching, referred to as process-product research, and from PBTE.

Thus, the diagnostic inservice training used in this study combines elements from the theories that have been discussed. In it, the teaching tasks of any instructional unit are taught by the teacher in an unspecified manner. The students are then tested by means of criterion-referenced tests on this unit. The tests are prepared so that they can be used by anyone teaching the unit. Test preparation, administration, and correction are conducted by someone other than the teacher to assure objectivity.

The graphed test results form the basis for the corrective feedback from the teacher-trainer. As each diagnostic inservice training cycle repeats itself over a period of time and results in more positive feedback, the reinforcement and subtle pressure to succeed becomes stronger, so that the teachers are more actively trying to improve their presentations to receive the positive reinforcement that the results and feedback offer. This improvement should result in higher reading achievement test scores.

Furthermore, as the teachers continue to improve their teaching performance they are less likely to have to repeat lessons or to teach partial lessons because the lessons were not grasped the first time. If: (a) the students can learn the material the first time it is presented, (b) the criterion-referenced test results confirm that the students are learning the material, and (c) the teachers can overcome presentation problems through suggestions given by the teacher-trainer, then the teachers in the treatment group should be able to cover more lessons than the teachers in the comparison group.

Method

This study was conducted in an urban school district that has poor pupil achievement, heavy bilingualism and low pupil SES. The study utilized an experimental pretest-posttest comparison group design. To minimize Hawthorne effects, all teachers were notified that they were part of the study. Forty-seven first grade teachers were clustered by school and then randomly assigned to the treatment group. Control for gross differences was achieved by matching the 14 schools by SES status and using the pretest scores as a covariate.

All teachers were instructed to use the district's specified language program daily. A series of criterion-referenced group tests, based on 10-lesson blocks of this program, were developed, administered and graded by the experimenter, who administered them monthly to the treatment group. The results were charted by skill and by student and were returned to each teacher by the teacher-trainer. The teacher-trainer, experienced in direct instruction principles,

met with each teacher on-site. She discussed the test results, modeled skills that were failed- recommended alternate teaching strategies, and suggested reteaching part and/or whole lessons as necessary.

The mean CAT, Level 11, NCE reading score for each class was used to measure pretest and posttest achievement. All teachers were surveyed at the end of the treatment period to determine which lesson they had reached in the language program.

To compare treatment and comparison groups in reading achievement, posttest mean NCE scores for each subtest of the CAT and Total Reading mean scores were compared by means of an ANCOVA, using the pretest CAT mean as a covariate. To compare treatment and comparison groups in the number of lessons presented, final lesson day in the language program was analyzed by means of a one-tailed t-test.

Ancillary analyses were conducted to explore relationships between the reading achievement test scores and the number of lessons covered in the language program and teacher demographic variables. After attrition, the final sample was composed of 20 teachers in the experimental group and 21 teachers in the comparison group.

Results

The California Achievement Test, Level 11, was administered to all classes in October and May by the classroom teachers as part of the district testing program. Scores of limited English proficient students were disaggregated and all data were analyzed by an independent scoring service. There were three subtests: Phonics, Reading Vocabulary, Reading Comprehension, and a Total Reading score.

Preliminary two-tailed t-tests showed no significant differences ($p > .05$) between the groups on any of the pretest CAT scores or teacher demographic variables.

Table 1 summarizes the pretest and posttest results for experimental and comparison groups on each subtest of the CAT.

Although the comparison group began with higher pretest results in every subtest, the experimental group scored higher in every area except Reading Vocabulary on the posttest, and made greater gains in every subtest. However, none of the results reached significance.

Tables 2-5 summarize the posttest mean NCE scores for Phonics, Reading Vocabulary, and Reading Comprehension subtests and Total Reading, that were compared by means of a series of ANCOVAs, using the pretest CAT mean as a covariate. Although the amount of variance explained by the models is significant in each subtest, significance stems from the covariate, which is the pretest score, rather than from the main effect, possibly due to the small N.

Table 1

Class Reading Achievement Test Results for California Achievement Test, Level 11, for English-Speaking Students

Test period	Subtest: Phonics NCE M	SD	NCE range	Subtest: Reading vocabulary NCE M	SD	NCE range
Experimental Group (N = 20)						
Pretest	36.6	9.7	23-70	38.5	9.2	26-65
Posttest	45.9	11.9	27-76	45.1	12.5	26-72
Gain	9.3			6.6		
Comparison group (N = 21)						
Pretest	40.5	7.4	31-56	40.8	8.6	29-59
Posttest	42.2	12.9	22-69	46.5	10.8	27-66
Gain	1.7			5.7		

Test period	Subtest: Reading comprehension NCE M	SD	NCE range	Total reading NCE M	SD	NCE range
Experimental Group (N = 20)						
Pretest	41.6	12.7	12-65	35.4	15.4	8-70
Posttest	45.9	13.3	24-77	45.7	13.2	24-77
Gain	4.3			10.3		
Comparison group (N = 21)						
Pretest	44.6	9.6	26-62	37.3	12.0	12-64
Posttest	45.8	10.7	25-66	45.1	11.8	28.67
Gain	1.2			7.8		

Table 2

Posttest Adjusted Analysis of Covariance of Class Reading Achievement, California Achievement Test, Level 11, Phonics Subtest

(N of classes = 41)

Source of variation	Sum of squares	df	s^2 M	F	Significance of F
Covariates	978.515	1	978.515	7.576	.009
Pretest phonics score	978.515	1	978.515	7.576	.009
Main effects	156.524	1	156.524	1.212	.278
Diagnostic inservice trng.	156.524	1	156.524	1.212	.278
Explained	1135.039	2	567.519	4.394	.019
Residual	4907.937	38	129.156		
Total	6042.976	40	151.074		

Table 3

Posttest Adjusted Analysis of Covariance of Class Reading Achievement, California Achievment Test, Level 11, Reading Vocabulary Subtest

(N of classes = 41)

Source of variation	Sum of squares	df	s^2 M	F	Significance of F
Covariates	1947.288	1	1947.288	21.920	.001
Pretest vocabulary score	1947.288	1	1947.288	21.920	.001
Main effects	1.307	1	1.307	.015	.904
Diagnostic inservice trng.	1.307	1	1.307	.015	.904
Explained	1948.595	2	974.297	10.967	.001
Residual	3375.844	38	88.838		
Total	5324.439	40	133.111		

Table 4

Posttest Adjusted Analysis of Covariance of Class Reading Achievement, California Achievement Test, Level 11, Reading Comprehension Subtest

(N of classes = 41)

Source of variation	Sum of squares	df	R^2	F	Significance of F
Covariates	799.201	1	799.201	6.270	.017
Pretest comprehen. score	799.201	1	799.201	6.270	.017
Main effects	18.870	1	18.870	.148	.703
Diagnostic inservice trng.	18.870	1	18.870	.148	.703
Explained	818.070	2	409.035	3.209	.052
Residual	4843.735	38	127.467		
Total	5661.805	40	141.545		

Table 5

Posttest Adjusted Analysis of Covariance of Class Reading Achievement, California Achievement Test, Level 11, Total Reading Scores

(N of classes = 41)

Source of variation	Sum of squares	df	R^2	F	Significance of F
Covariates	1817.045	1	1817.045	16.248	.001
Pretest total reading score	1817.045	1	1817.045	16.248	.001
Main effects	25.082	1	25.082	.224	.639
Diagnostic inservice trng.	25.082	1	25.082	.224	.639
Explained	1842.127	2	921.063	8.236	.001
Residual	4249.629	38	111.832		
Total	6091.756	40	152.294		

Table 6 summarizes the results for experimental and comparison groups on the number of lesson days covered in the language program. A one-tailed t-test was utilized because it was hypothesized that the group receiving the diagnostic training would cover more lessons than the comparison group.

The mean for the experimental group was 56.7, as opposed to 34 for the comparison group. Separate variance estimates were utilized to obtain the t-value. The t-value was 1.94 with 28.72 degrees of freedom, with $p = .032$. The experimental teachers, who received the diagnostic inservice training, covered significantly more lesson days in the language program than did the comparison group teachers.

Table 6

Comparison of Number of Lessons Covered

						Separate variance estimate	
Group	N	M	SD	Standard error	t value	df	1-tailed prob.
Experim.	20	56.7	22.73	5.08	1.94	28.72	.032
Comprsn.	21	34.0	48.39	10.56			

Ancillary analyses were performed to determine whether there were relationships between the number of lesson days covered, posttest CAT scores and demographic data for experimental and comparison groups. A series of Pearson r correlations, using a two-tailed test of significance were performed on the data. These data are summarized in Table 7.

There were no significant correlations between lesson day in program scores and demographic variables for either group ($p > .05$). The direction of the correlations seemed to show that the older, more educated and more experienced teachers in the comparison group were the ones who tended to make less use of the language program, but instead followed their own dictates. The younger, less experienced teachers tended to follow the language program, but without the diagnostic inservice training, there was no correlation between the amount they taught in the program and the subsequent CAT scores.

Table 7

Pearson Correlations of Number of Lessons Covered (NLC) for Experimental and Comparison Groups

Variable	Group: Experimental (N = 20)		Group: Comparison (N = 21)	
	r	p	r	p
Posttest phonics	.465	.039	.273	ns
Posttest rdg. vocabulary	.479	.033	-.060	ns
Posttest rdg. comprhnsn.	.542	.014	-.229	ns
Posttest total rdg.	.508	.022	-.223	ns
R +	-.075	ns	.214	ns
R -	.224	ns	-.071	ns
Teacher age	.217	ns	-.130	ns
Teacher experience	.293	ns	-.068	ns
Teacher education	.360	ns	-.005	ns

Note. Two-tailed test of significance.

For the experimental group, covering more lessons in the language program, together with the diagnostic inservice training, resulted in a significant positive correlation between lesson day in program and the posttest CAT scores on every subtest. The more the experimental group teachers covered in the language program, the higher the resulting CAT scores in Phonics ($p = .03^{o}$), Reading Vocabulary ($p=.033$), Reading Comprehension ($p=.014$), and Total Reading ($p = .022$).

Discussion

This study was begun with the knowledge that a Type II error was a strong possibility, due to the small sample size. This chance was intensified with a loss of some of the teachers due to attrition, which was partially due to the district's belated decision to lower class size in grade 1, causing an inordinate number of new teachers to be assigned to that grade, not all of whom remained. Because of this decision, the experiment did not begin until November. Therefore, diagnostic inservice training did not begin until December and could only continue until early May, when the posttest was administered. Given the positive direction of the reading achievement results obtained by the experimental group, it is possible that a longer experimental time period would have been more successful, particularly one that extended through a second school year.

Another problem involved the agreement between the experimenter and the participating teachers, which was that no information would be shared with the school supervisors. This was necessary to obtain the voluntary participation of the teachers. The advantages were that the experimental group teachers were able to express their problems freely and to solicit advice from the teacher-trainer without fear of censure. They were able to admit not having taught the language program as often as they should have. This freedom is an advantage, as teachers are more accepting of advice that is voluntary. The disadvantage of this approach is that the teachers cannot be monitored by the school administration on their use of an instructional program.

If this diagnostic inservice training procedure were adopted as part of the normal school supervisory program, it might lose some of the advantages gained from any voluntary program. However, there would be a gain in terms of implementation. It would also give school supervisors and staff development personnel a new perspective from which to assist teachers to increase their effectiveness.

Because of the chance that a Type II error has occurred, it is only possible to state that it could not be proven that diagnostic inservice could have a significant impact on these scores. Indeed, as every gain in scores favored the experimental group, it is entirely possible that a larger population size, longer duration of the experiment, or more intense treatment might have been successful.

In spite of the results of the studies of Anderson, Evertson and Brophy (1979) and Good and Grouws (1979), showing significant gains with only very minimal interventions, a later experiment by Colodarci and Gage (1984) failed to replicate these results. It is interesting to note that the Anderson et al experiment involved reading in white, middle-class schools. The Good and Grouws experiment dealt with mathematics. The Colodarci and Gage experiment involved reading and math in lower-class black classes. The present experiment had a student population that was mostly Hispanic, in a lower-class urban area.

It is possible that the type of teacher training needed to raise reading achievement levels is compounded by the impoverished, language-deficient backgrounds present in many lower-class students. Perhaps minimal interventions can only be successful in raising reading levels for middle-class students.

In the present study, the most consistent indicator of posttest performance was the pretest score. This conforms, unfortunately, to many of the conclusions in the Coleman report (1966), which have been difficult for educators to overcome, particularly with lower- SES and language-impoverished populations.

Ancillary multiple regression analyses that were performed to discover the relationships between teacher status items, class reading achievement test scores and diagnostic inservice training show that pretest scores, which accounted for between 14% and 30% of the variance (depending on the subtest), and teacher experience, which accounted for an additional 8% to 23% of the variance, were both highly significant in determining the class reading achievment test results.

However, if these multiple regression analyses are compared with an analysis that used the number of lesson days covered as the dependent variable, in which there were no pretest scores, and all classes started at the same lesson day, without prior assumptions, the variable of teacher experience was negated.

The implications of these analyses again point to the need to overcome initial differences caused by both pretest scores and teacher experience. If teacher experience cannot be controlled for, then assigning teachers specific curricula to follow, and giving them assistance, such as diagnostic inservice training, should help overcome these prior conditions. This is particularly important for many schools located in lower socio-economic areas where teacher experience is low and turnover is very high. It points to the need for more research, which might vary the amount of diagnostic inservice training according to the teacher's experience.

The two-tailed correlational analyses between the number of lessons covered and posttest reading achievement scores showed striking differences between the experimental and comparison groups. If there had been an inherent relationship between the number of lessons covered in the language program and the standardized reading achievement results, then that relationship would have been evidenced by significant correlations between these variables for the comparison group as well. Teachers in the comparison group who failed to use the language program, or used it to a greater extent, should have obtained commensurate scores on the reading achievment tests. The lack of this relationship for the comparison group points to the diagnostic inservice training as the mitigating variable in obtaining the significant relationship. It justifies treating reading achievement scores and number of lessons covered as two separate dependent variables, as was assumed for this study. The relationship between the number of lessons covered and the reading achievement posttest scores was reached only because of the addition of the diagnostic inservice training.

References

Acheson, K.A., & Gall, M.D. (1980). Techniques in the clinical supervision of teachers: Preservice and inservice applications. New York: Longman.

Anderson, C., Evertson, C., &Brophy, J. (1979). An experimental study of effective teaching in first grade reading groups. Elementary School Journal, 79, 193-223.

Bloom, B.S. (1976). Human characteristics and school learning. New York: McGraw-Hill.

Coleman, J.S., Campbell, E.Q., Hobson, D.J., McPartland, J., Weinfeld, F.D., & York, R.L. (1966). Equality of educational opportunity. Washington, D.C.: United States Government Printing Office.

Colodarci, T., & Gage, N.L. (1984). Effects of a minimal intervention on teacher behavior and student achievement. American Educational Research Journal, 21 (3), 539-555.

Gage, N.L., & Winne, P.H. (1975). Performance-based teacher education. In K. Ryan (Ed.), Teacher education: The seventy-fourth yearbook of the National Society for the Study of Education, Part 2. Chicago: The National Society for the Study of Education, 146-172.

Good, T.L., & Grouws, D.A. (1979). The Missouri mathematics effectiveness project. Journal of Educational Psychology, 71, 355-362.

Joyce, B.R., & Showers, B. (1981). Transfer of training: The contribution of "coaching," Journal of Education, 163 (2), 163-172.

Meyer, L.A., Gersten, R.M., & Gutkin, J. (1983) Direct Instruction: A Project Follow Through success story in an inner-city school. Elementary School Journal, 84 (2), 241-252.